2008-01-2039

Operator Performance Evaluation of Fault Management Interfaces for Next-Generation Spacecraft

Miwa Hayashi, Ph.D., Ujwala Ravinder, M.S.
San Jose State University & NASA Ames Research Center, Moffett Field, CA

Brent Beutter, Ph.D., Robert S. McCann, Ph.D., Lilly Spirkovska, Ph.D.
NASA Ames Research Center, Moffett Field, CA

Fritz Renema, M.S.
Perot Systems Government Services & NASA Ames Research Center, Moffett Field, CA

ABSTRACT

In the cockpit of NASA's next generation spacecraft, most vehicle command will be performed via electronic interfaces instead of hard cockpit switches. Checklists will be also displayed and completed on electronic procedure viewers rather than on paper. Transitioning to electronic cockpit interfaces opens up opportunities for more automated assistance, including automated root-cause diagnosis capability. The paper reports an empirical study evaluating two potential concepts for fault management interfaces incorporating two different levels of automation. The operator performance benefits produced by automation were assessed. Also, some design recommendations for spacecraft fault management interfaces are discussed.

INTRODUCTION

During the time-critical, dynamic flight phases of launch, ascent, and entry, the spacecraft crew must monitor the systems' parameters on the cockpit interfaces carefully, so that they can react quickly when any problem occurs. Numerous sensors are installed throughout the spacecraft, and their readings are presented to the crew through the cockpit displays. This huge amount of information is difficult for any single human operator to thoroughly monitor without automated assistance.

In the Space Shuttle cockpit, a relatively simple Caution & Warning (C&W) system assists the crew in detecting system anomalies by issuing audio and visual alarms whenever a sensor reading exceeds its predefined limits. At the same time, a brief fault message is displayed for each out-of-limit sensor reading and associated parameters on the status displays turn to yellow (if caution) or red (if warning). These events will be collectively called *C&W events* in this paper. A major problem in such a C&W system is that the failure of one critical item often causes many additional downstream components dependent on it to also cease functioning properly. This produces C&W events not only for the root cause, but also for the failures of each of the downstream components and connected systems dependent upon it, often resulting in a flood of C&W events [1]. It is then up to the crew to interpret the multitude of C&W events, figure out which one is the root cause, and work on mitigating the root-cause failure. Furthermore, after identifying the potential root cause, the crew must manually 1) locate the appropriate section of the paper checklist, 2) search for and find the appropriate checklist, and 3) execute the relevant checklist steps to perform the necessary system reconfiguration. These operations can be slow, inefficient, attention demanding, and error prone.

Transitioning from the Shuttle to the next-generation spacecraft, the Crew Exploration Vehicle (CEV) Orion, provides a great opportunity to introduce more automation technologies to the crew's fault management processes. The CEV will have much less interior volume than the Shuttle, and with little interior room for physical switch panels, virtually all vehicle commanding will take place through soft (electronic) switches provided on the displays. Also, paper checklists will be eliminated and replaced with an electronic version. The electronic checklists potentially offer opportunities for simple automation assistance not available on paper checklists, such as automatically highlighting the currently active line, checking off completed lines, skipping lines that belong to inapplicable branches of conditionals, and so forth. Furthermore, the electronic checklists can be integrated with the status displays and the electronic switch panels. For instance, the electronic checklists can automatically highlight the parameter or the switch associated with the current checklist line, thus quickly drawing the crew's attention to the right place.

This study examined some of the potential design options of advanced fault management interfaces, namely the Advanced Caution and Warning System (ACAWS) displays, which utilize an integrated electronic checklist. One of the major design options of interest was whether or not to include the automated root-cause diagnosis capability. It was reported that commercial airline pilots have been expressing the desire for real-time, onboard automated root-cause diagnosis capability to help save precious cognitive resources and time under high-workload, potentially life-threatening situations [2]. The same reasoning applies to spacecraft operations. The root cause information also can be utilized in the electronic checklists to automatically call up the proper checklist, saving the crew's time and effort to do so manually and reducing the probability of the crew selecting a wrong checklist. The downside of including such capability is the additional vehicle hardware and weight required to provide the increased computational resources. The development and maintenance of complex root-cause diagnosis algorithms also will be costly and time-consuming. In addition, there are some human factors concerns. For instance, what if the computer chooses an incorrect root-cause diagnosis? Will the crew be able to correctly identify the root cause, manually override the computer's recommendation, and call up the correct checklist?

Another major design option considered in this study was whether the interface should present all the detailed system information or just a summary of it suppressing the details that the computer thinks less important to the crew (the suppressed details are always available when the crew takes an action to call them up). Another design option examined was whether the switches should be accessed via a virtual switch panel with switch icons resembling hardware switches, or the switches should be integrated with the system line-diagram and directly accessible via the switch symbols within the diagram. In the former option, the virtual switch panel is presented over the line diagram due to the limited display space, while in the latter option, the switch symbols are in the line diagram, and thus, the line diagram is always visible, enabling the crew to immediately observe the effects of a switch throw on the system.

This paper reports the results of a simulator study examining the impacts of several ACAWS display design options on operator performance. Two prototype ACAWS displays, called Elsie and Besi (not acronyms), were designed for fault management of the Electrical Power System (EPS) and Environmental Control and Life Support System (ECLSS). The EPS and ECLSS used in the study were simplified versions of those of the Shuttle. The two systems are tightly interconnected because EPS supplies electrical power to the ECLSS loads (e.g., fans, pumps). Therefore, when an ECLSS component indicated an out-of-limit sensor reading and generated C&W events, the root-cause could lay in either the ECLSS itself or in the EPS. The participants' task was to identify the root cause and perform the proper procedure using the ACAWS displays. Elsie displayed detailed EPS/ECLSS information, had virtual switch panels separate from the system line-diagram, and did not provide automated root-cause diagnosis. Besi summarized EPS/ECLSS information, had switch symbols integrated in the system line-diagram, and provided automated root-cause diagnosis. In short, Elsie was somewhat similar to current Shuttle displays with the exception of electronic switches and checklists, while Besi had more automation assistance. The EPS-related displays and ECLSS display for Elsie and Besi were developed for this study, and the other displays were similar to those of the Space Shuttle Cockpit Avionics Upgrade (CAU) display suite [3]. (CAU is sometimes pronounced as "cow;" thus, the ACAWS displays used in this study were named after cow names.)

In addition to the EPS/ECLSS fault management tasks, the participants were asked to concurrently monitor the flight variables on the Primary Flight Display (PFD) to measure their ability to divide their attention between the fault-management and flight-monitoring tasks. At different times, one of the five major flight variables of the CEV changed its color from white to yellow, and then to red. The participants were asked to respond to the color change by touching the flight variable that had the color change. If the participant missed a color change, that indicated that the EPS/ECLSS task had caused attention tunneling (i.e., a tendency to focus very hard on only one thing, ignoring surrounding information, often observed under high-stress situations [4]).

This study analyzed and compared the participants' EPS/ECLSS fault-management performance accuracy, fault resolution times, PFD task misses, eye movements, subjective workload scores, and subjective preference ratings between Elsie and Besi. The primary goals of this study were to provide empirical data of the performance benefits provided by Besi's automation, as well as to provide generic design recommendations for the ACAWS displays.

METHODS

SIMULATOR - The CEV cockpit simulator consisted of an HP XW9300 workstation with two dual-core 2.4GHz AMD Opteron processors, 2GB of system memory, and an NVIDIA Quadro FX3450 video card with 256MB of memory running Windows XP Professional. The workstation was used for PFD and ACAWS display graphics generation, data collection, and sound control. The video card drove two 20-inch touch-sensitive Apple Cinema displays, one of which was used by the experimenter and the other by the participant. Figure 1 shows the display used by the participants (the display unit on the right was not used). The top half showed the PFD and the bottom half showed the ACAWS display. The experimenter's terminal was located outside the simulator room. The simulator included a speaker system to generate background engine noise and issue auditory alarms.

Figure 1: CEV Simulator

Another HP XW9300 with a similar configuration was dedicated to real-time computation of the EPS and ECLSS parameters with Matlab (version 7.4.0) and Simulink (version 6.6.1). In Besi trials, this workstation also ran automated model-based root-cause diagnosis software, Hybrid Diagnostic Engine ("HyDE") [5]. Simulated CEV flight parameters were computed real-time on the third workstation, a Sun Fire V40z Server, with two dual-core 2.4Ghz AMD Opteron processors and 16GB of system memory. The flight dynamics were based on a CEV flight model developed at NASA Johnson Space Center.

The simulator was also equipped with a head-mounted eye tracker (ISCAN ETL-500, ISCAN, Inc., Burlington, MA) integrated with a head tracker (FasTRAK, Polhemus, Colchester, VT) to measure the participants' gaze locations. The eye-tracking system was mounted on a baseball cap as shown in Figure 1. The display area subtended a visual angle of 30.2° horizontally and 47.2° vertically at a viewing distance of 20 inches. The sampling rate of the eye-movement data was approximately 60 Hz, and the spatial resolution was approximately 0.5 inches.

EPS/ECLSS MODEL - Figure 2 shows our EPS model. It contained three batteries (A, B, and C) capable of supplying power to the ECLSS loads connected to the two load banks (L1-L7 on Load Bank A, and L1-L8 on Load Bank B). Switches are shown as pairs of two circles with a short line in between (their names are also underlined), and circuit breakers as coaxial double circles. Figure 2 illustrates the initial configuration of the EPS at the beginning of each trial, where Battery A powered Load Bank A, and Battery B powered Load Bank B. Battery C was a backup battery. For each load bank, L1 through L6 were AC Loads, and L7 and L8 were DC Loads. Loads L1 and L2 on each load bank were the primary critical loads necessary for crew survival, loads L3 and L4 were non-critical loads, and loads L5 and L6 were backup critical loads for the other load bank's critical loads. Load L7 on Load Bank A was a primary critical DC load, and load L8 on Load Bank B was the backup for Load Bank A's L7. In the initial configuration, the load switches for the backup loads were off (open).

The participants were instructed to resolve any malfunctions by calling up the appropriate checklist (i.e.,

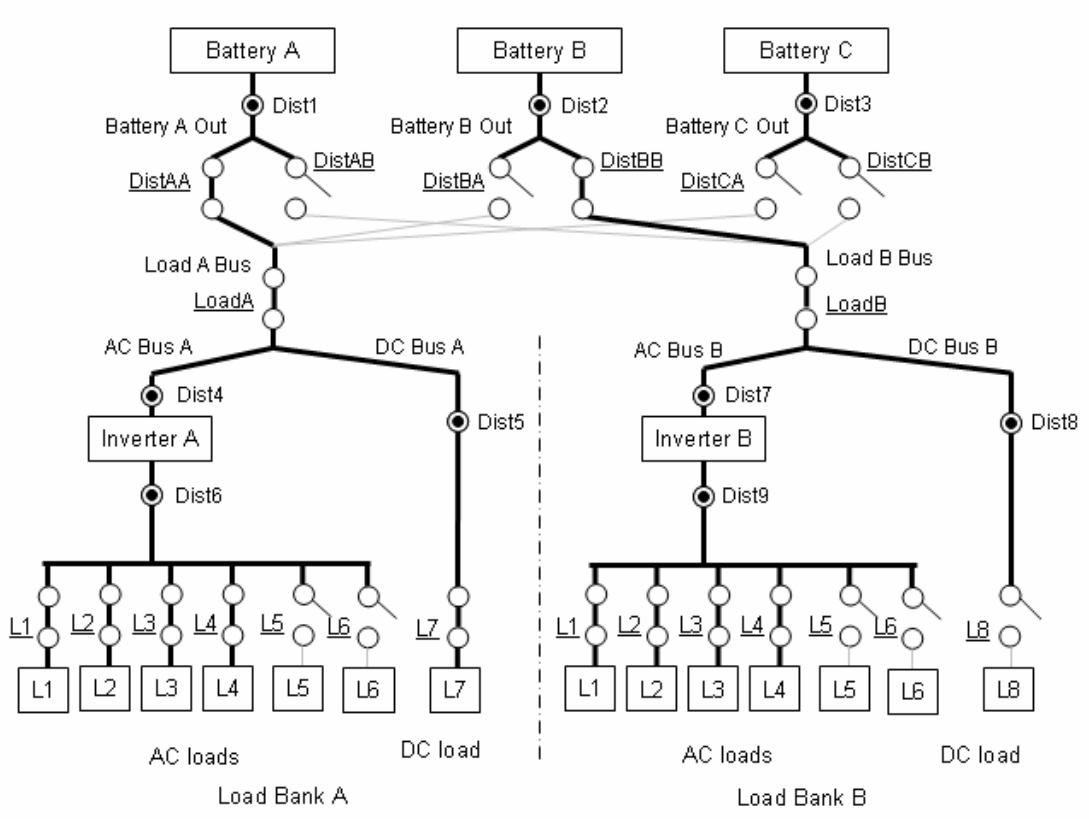


Figure 2: Electrical Power System (EPS) in initial configuration

performing from memory was not allowed). The process to select the appropriate checklist was different between Elsie and Besi, and described in the ACAWS Displays section. Once the checklist was opened, the procedure usually started with diagnosing the problem by troubleshooting and/or cross-checking related system parameters. After the diagnosis steps were completed and the problem was verified, the checklist instructed recovery steps to reconfigure the system (e.g., activate proper backups). For instance, for a failed-open switch (i.e., stuck at an open position), the first steps in the checklist consist of a diagnosis of the problem by cycling the switch. If cycling restores the switch, the problem is resolved. If not and the switch is verified to be un-restorable, the checklist instructs additional switch throws to activate the appropriate backup components.

Participants were instructed to follow these rules for switches:

- Before cycling a switch, all the ECLSS loads connected to the switch must be turned off. This prevents damage to the loads by a high inrush current later.
- One load bank should not be powered by more than one battery at any moment.
- *The load-shedding rules:* To avoid exceeding the maximum power capacity limit of a single battery, one or two non-critical loads (L3 or L4) must be turned off before turning on a backup critical load (L5 or L6).

These rules are complex, but strictly following the checklist procedures insures that these rules are satisfied, except for the load-shedding rules. Because the number of the primary critical loads lost may be different in each case (especially when there was a previous malfunction that disabled some of the primary critical loads), the checklist did not specify how many non-critical loads had to be shed. Thus, the participants needed to determine which non-critical loads had to be shed each time.

ACAWS DISPLAYS - The bottom half of the Apple Cinema display presented the ACAWS display (Figure 1). The participants interacted with the ACAWS display through a commercial hand controller manufactured by Nostromo (Figure 3). The rationale of using this type of hand controller was to simulate a dynamic flight phase situation (e.g., ascent), where large G-forces and vibration require the crews' arm to be restrained. This

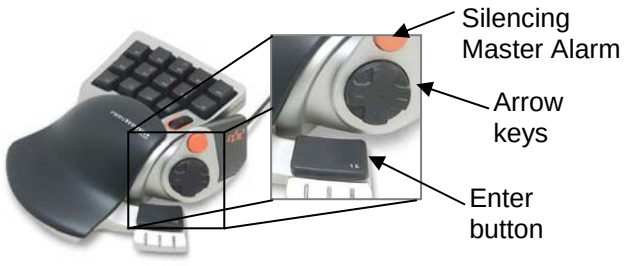


Figure 3: Nostromo hand controller

requires the use of a hand controller installed near the arm rest to interact with the display. The arrow keys were used to move the cursor among edge keys and other selectable items in the display, and the Enter button was used to select the item. Pressing the orange round button silenced the Master Alarm sound.

Elsie – Figure 4 shows Elsie. In this figure, the main area (the large square area) shows the EPS Summary (EPS Sum) display. The checklist area is on the right side, and the fault message area is below the main area. The red edge key at the lower left corner is the Master Alarm. As mentioned earlier, Elsie presented detailed system information to the crew. The information was organized by several displays within Elsie: the EPS Sum showed the EPS diagram (Figure 4), EPS Main showed all the EPS parameters in a table format (Figure 5), and EPS Loads showed the status of the load switches in a table format (not shown). The ECLSS parameters were presented on the ECLSS display (not shown). All of these displays, as well as the EPS switch panels (Figure 6 shows an example), were accessible via the edge keys. All of these displays appeared in the main area.

When there was no system malfunction to work on, such as during the beginning of the trial, the participants were instructed to bring up the Fault Summary (Fault Sum) display (Figure 7) in the main area to monitor the overall health of the EPS and ECLSS. Once any out-of-limit sensor values were detected, the C&W system issued the C&W events, including a visual alarm (red Master Alarm), auditory alarms, color changes of the affected parameters on the Fault Sum display (yellow for caution, red for warning), and fault messages. The participants used the Fault Sum display information to judge which system, EPS or ECLSS, was likely to contain the root cause. It was also recommended for the participants to go check the EPS Sum display (if the suspected component was in EPS) or the ECLSS display (if it was in ECLSS) to quickly assess which specific component likely had the root-cause failure.

Then, the participant checked the fault messages displayed in the fault message area (lower part of Elsie). This area displayed only the first five messages. If there were more, the participant needed to go to the Fault Log display to view the entire list (Figure 8). The Fault Log had three pages, each of which could list up to eighteen messages, or 54 in total, in chronological order. To determine which message was associated with the root cause, the participants were instructed to look for, first, any “switch mismatch” fault message (i.e., the commanded position and sensed position of the switch did not match). If there was no such message, then, second, they looked for the volts-related message coming from the most upstream part of the system (for example, the “Load A volts low” message is upstream of the “DC Bus A volts low”; see Figure 2). In Elsie, the participant must go through the fault messages and find the one that was associated with the potential root cause, because that fault message was also the title of the checklist to be selected.

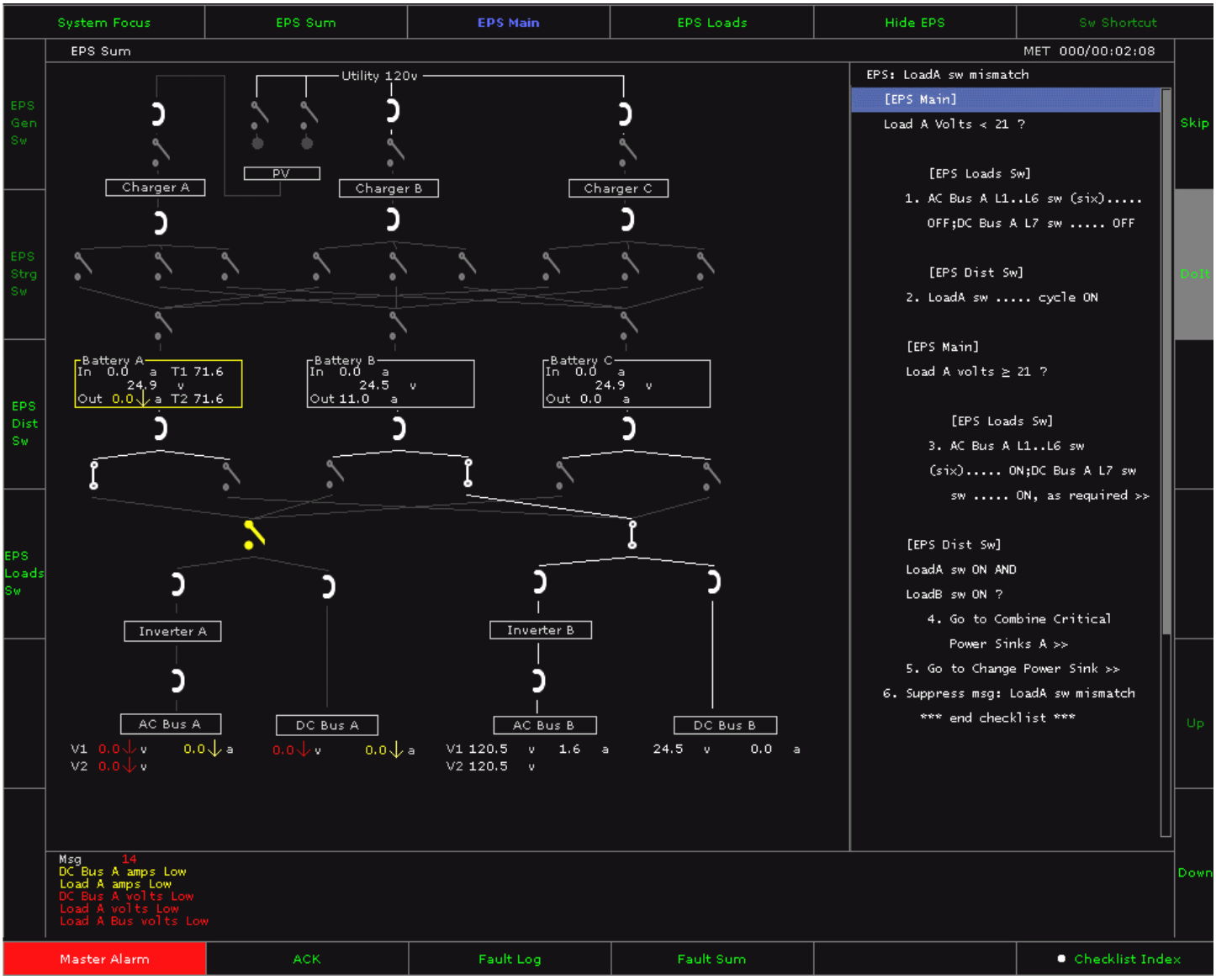


Figure 4: Elsie. The main area is showing the EPS Sum display.

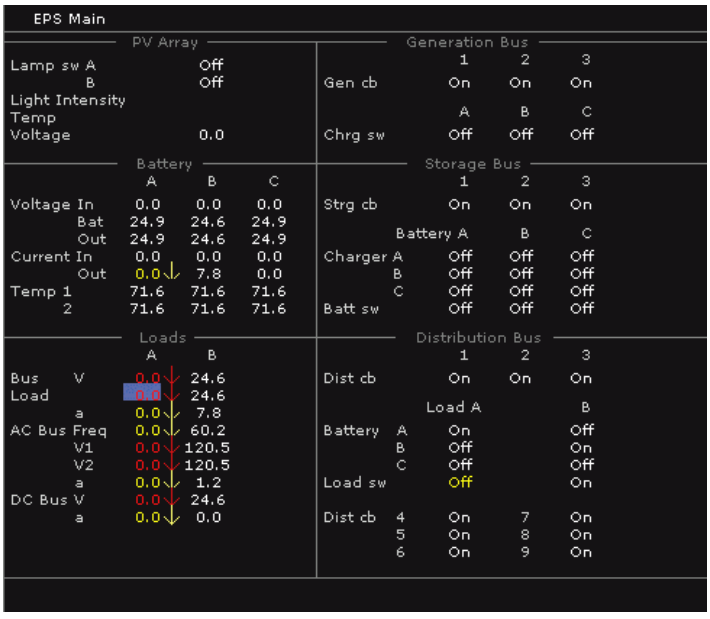


Figure 5: Elsie - EPS Main

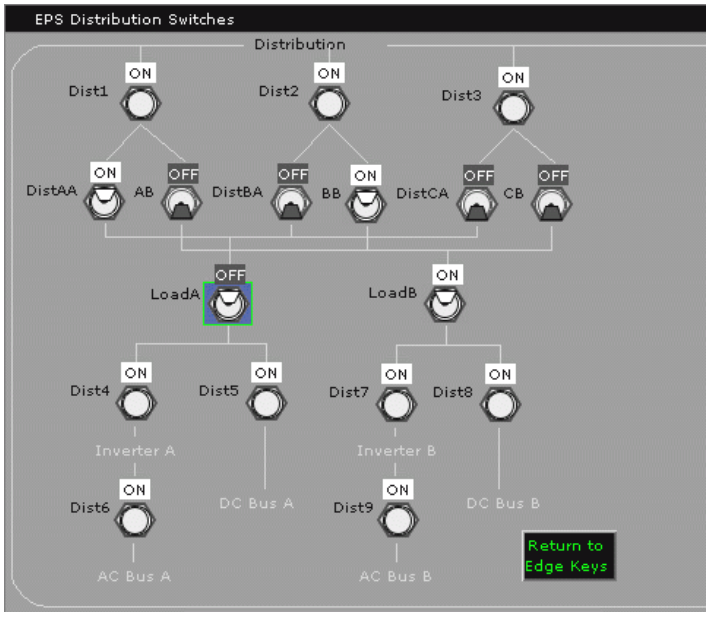


Figure 6: Elsie – One of EPS switch panels

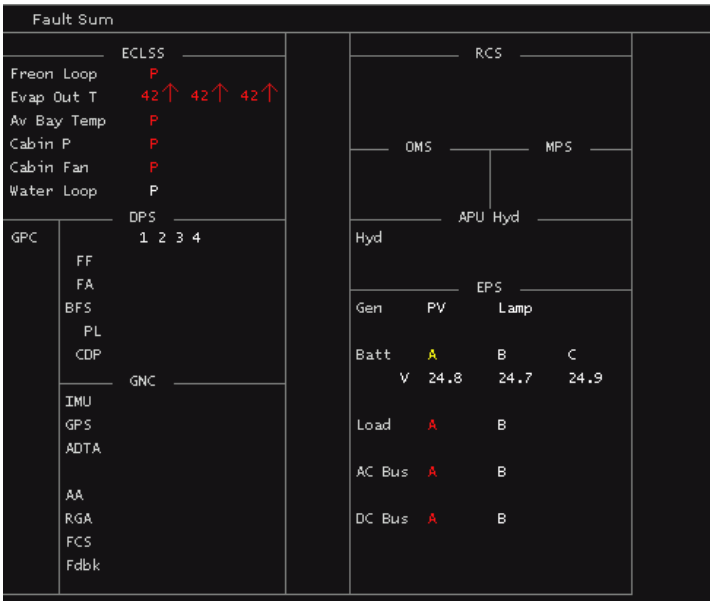


Figure 7: Elsie – Fault Sum

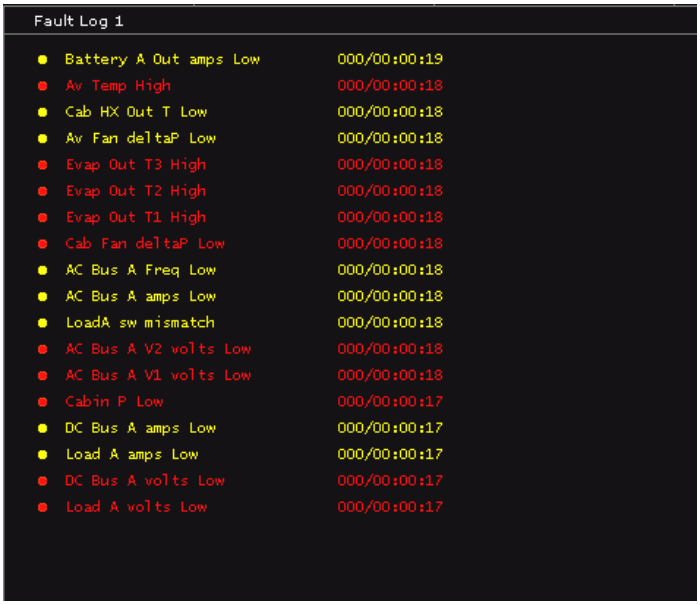


Figure 8: Elsie - Fault Log

After identifying which checklist to go to, the participant selected the *Checklist Index* edge key at the lower right corner to bring up a list of checklists in the checklist area. Then, the participant manually navigated to the target checklist through the list(s). Once the checklist was opened, the participants used the set of edge keys on the right side to navigate among the lines. As mentioned before, the ACAWS's electronic checklist provided the operator some automation assistance not available on the traditional paper checklists. For instance, if the checklist line shows the display name in square brackets (e.g., [EPS Main]), that means the procedure requires the display to be brought up. In the ACAWS checklist, selecting the *Do It* edge key automatically brings up the display in the main area. Also, the system parameters or the switch icons associated with the current checklist line were automatically highlighted in blue, making it easier for the participants to find them (see Figures 5 and 6 for examples). The cursor could be moved to the highlighted switch icon via the *Sw Shortcut* edge key at the upper right corner, and moved back to the edge keys via *Return to Edge Keys* icon on the switch panel.

Some procedures may require opening separate checklists for the diagnosis and recovery processes. In that case, the participants had to manually navigate to the checklist each time. When the checklist procedures were complete and the malfunctions were resolved, the participant brought up the Fault Sum display to confirm that the problem was resolved and continue monitoring the EPS/ECLSS status.

Besi - Figure 9 shows Besi. In this figure, the main area is showing the EPS display. In Besi, all pieces of the EPS-related information, including the switch panels, were consolidated onto the single EPS display. When the checklist called for a switch throw, the participant moved the cursor to the switch symbol within the line diagram via the *Connect* edge key, and toggled the switch state by pressing the *Enter* button.

The EPS display in Besi provided graphical representations for voltage (V), current (A), and battery temperature (T) for all active elements. The height of the fills indicated the sensor readings, and their limits were marked as red ticks. No numerical data were presented on Besi by default. As mentioned before, Besi provides summarized information to the operator with the details still accessible by command. If a participant wished to see the numerical values, he/she could call them up by selecting the *View* edge key at the upper left corner and then selecting the *#s: all* option in the pull-down menu. Then the numbers appeared under the corresponding graphics. Also, by default, Besi's EPS display automatically suppressed the lines connecting the inactive components to reduce display clutter. The participant could see the suppressed lines by, again, selecting the *View* edge key, and then selecting the *conn: all* option.

The space below the main area was used to present three kinds of information rather than one in Besi: from left to right, system status, root cause list, and fault messages. The matrix in the system status area stayed dark gray if no out-of-limit sensor readings were detected. When any abnormal readings were detected, the name of the system that had a problem showed up in yellow (caution) or red (warning) and the numbers of caution and warning messages were displayed. Also, moving the cursor to the SUMM, EPS, or ECLSS cell in the system status matrix and selecting it brought up the Fault Sum, EPS, or ECLSS display, respectively. The Fault Sum and ECLSS displays were the same in Elsie and Besi. The root-cause list area presented the root cause computed by HyDE. The fault message area functioned in the same way as it did in Elsie. The Fault Log displays were also accessed via the edge key just like in Elsie. The Master Alarm edge key was shown at the top of Besi.

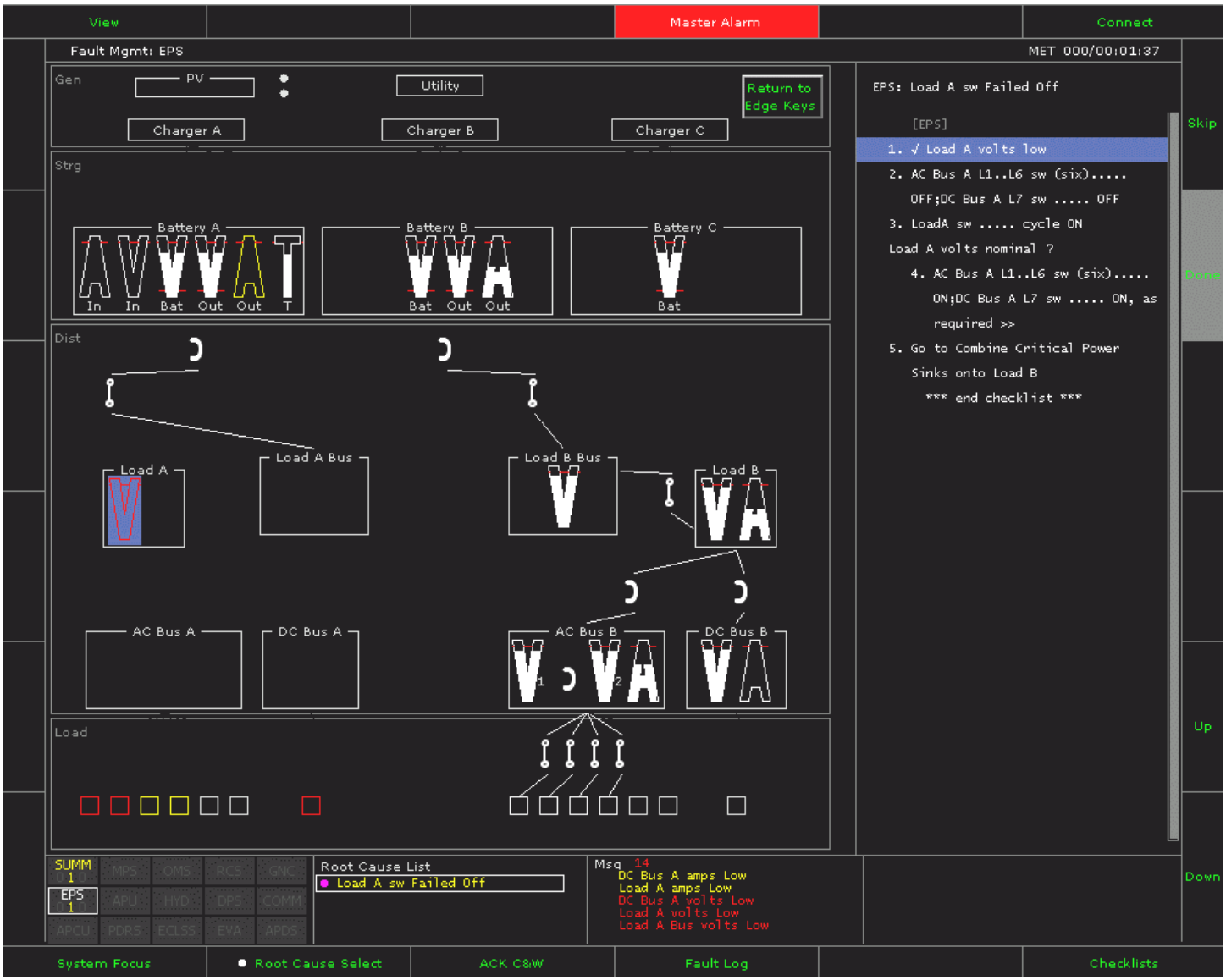


Figure 9: Besi.

As in Elsie trials, the participants were instructed to monitor the Fault Sum display whenever they were not working on any malfunction, including the beginning of the trial. When any out-of-limit sensor readings were detected, that issued the color changes of the affected parameters on the Fault Sum display (yellow for caution, red for warning) and the fault messages in the fault message area. The Master Alarm was not issued at this point yet. It usually takes five to seven seconds for HyDE to complete the root-cause diagnosis. While waiting, the participants could check the Fault Sum display, and then either the EPS or the ECLSS display as appropriate to examine which specific component may have failed. When HyDE completed the diagnosis, the root cause was displayed in the root-cause list area and the Master Alarm was issued to draw the participant's attention.

In Besi, the participants did not have to manually navigate to the checklist. Instead, they selected the *Root Cause Select* edge key to transfer the cursor into the root-cause list area, and selected the root cause that they wanted to work on. This brought up the appropriate checklist in the checklist area automatically. Notice that looking at the Fault Log display was no longer required in Besi, though the participants always had access to it.

The navigation within the checklist was done in an analogous way as in the Elsie checklist, using the edge keys on the right-hand side. Note that the checklists for the same root cause may be different between Elsie and Besi, because in Besi, HyDE had already performed all of the parameter cross-checking. Thus, instead of lengthy diagnosing steps typically seen in the Elsie checklists, the Besi checklists often contained only a single verification step (e.g., "√ Load A volts low," where "√" is read as *verify*) just to make sure that HyDE's computation was consistent with the situation. As in Elsie trials, some malfunctions may require using separate checklists for the diagnosis (or verification) and recovery processes. In Besi, however, when a checklist line requires going to another checklist, the participant can automatically jump to the next checklist by selecting the *Do It* edge key.

PFD TASK - The PFD used in this study included five flight parameters, i.e., altitude, velocity, G-meter, vehicle position, and thrust (see Figure 10). Each trial started with simulated liftoff of the ascent track, and every 20 seconds (on average) the interior of the box surrounding one of the flight parameters changed color from white to yellow, and after five seconds to red. The red color stayed for an additional five seconds and then turned back to white. When participants noticed a parameter box had changed color, they were instructed to, as quickly as possible, reach up and touch the indicator directly while calling out its name (for example, if the G-meter box changed color, they would call out "G-meter"). Touching the colored parameter returned its color back to white. In addition, the participants were asked to perform a nominal CEV call out when they detected the transition from the first stage to the second stage (by saying "Stage 2") and the transition from abort Mode 1 to Mode 2 (by saying "Mode 2").

PARTICIPANTS - Eight participants, all instrument-rated pilots, were recruited for the study. Instrument-rated pilots were specifically chosen because of their demonstrated ability to share attentional focus among multiple information sources. The participants included seven males and one female, and their ages ranged from 24 to 54 (average of 37.5). Their total flight times ranged from 230 to 21000 hours, and the instrument-flight times ranged from 68 to 2000 hours. All participants were right-handed and had normal or corrected vision (i.e., 20/40 or better).

TRAINING - Each participant received approximately twelve hours of training, including four hours of reading assignments at home, four hours of classroom lecture, and four hours of hands-on practice using a laptop-computer-based trainer. In addition, each participant had to take a final exam immediately prior to their first data-collection session and pass it in order to proceed to the data collection. The final exam served as a way to screen for any participants whose understanding was insufficient for adequately performing the fault management task.

DATA COLLECTION - Data collection was split into two sessions. Each session consisted of seven simulated ascents with one display format suite (either Elsie or Besi). Half of the participants completed seven Elsie trials in the first session followed by seven Besi trials in the second session. For the other half, the session order was reversed. Scenario orders were counterbalanced across participants.

Table 1 lists the fourteen malfunction scenarios used. Scenario pairs #5 and #6, #7 and #8, #9 and #10, and #13 and #14 were symmetric, and used in the different ACAWS displays within a participant. For instance, a participant who was assigned scenario #5 for Elsie was assigned scenario #6 for Besi, or vice versa. Using symmetric scenarios for Elsie and Besi within the same participant enabled the experimenters to compare Elsie and Besi performance directly while minimizing potential learning effects. Scenarios #11 and #12 were identical. Scenarios #3, #4, and #11 through #14 contained multiple malfunctions. In these scenarios, the second malfunction occurred 55 to 90 seconds after the occurrence of the first malfunction so that the participant was still working on the first malfunction when the second one occurred.

Scenarios #1 through #4 formed one group that provides 2 x 2 conditions, real failure vs. sensor failure and single malfunction (lower workload) vs. multiple malfunctions (higher workload). The switch sensor failures in #2 and

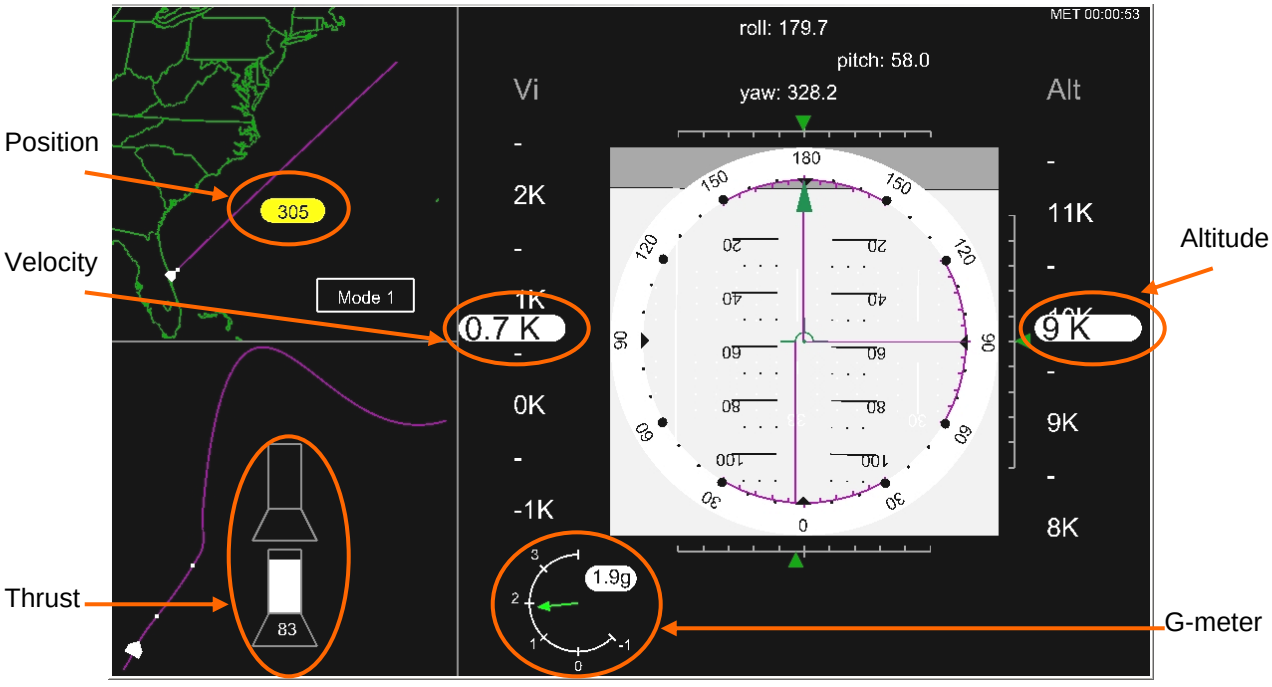


Figure 10: PFD

#3 generated a false alarm of switch failure when the switch was actually functioning correctly. For the Besi trials, the sensor failure also caused the HyDE to misdiagnose and generate a root cause that did not exist. (The original HyDE algorithms could distinguish sensor failures from actual component failures, but for this study, HyDE was intentionally modified to generate a false root cause in these particular cases.) Due to the different experiment designs, scenarios #1 through #4 were analyzed separately from the rest of the scenarios.

Table 1: Malfunction Scenarios

Sc #	Malfunction(s)
1	A/L1 switch failed open
2	B/L1 switch failed open (false alarm)
3	1) Load B switch failed open (restorable) 2) A/L2 switch failed open (false alarm)
4	1) Load A switch failed open (restorable) 2) B/L2 switch failed open
5	DistAA switch failed open (restorable)
6	DistBB switch failed open (restorable)
7	Battery A volts low
8	Battery B volts low
9	Inverter A failure
10	Inverter B failure
11	1) Inverter A failure 2) Battery A volts low
12	Same as 11
13	1) Battery A volts low 2) Battery B volts low
14	1) Battery B volts low 2) Battery A volts low

Each trial started from liftoff and ended at eight, ten, or twelve minutes, whichever came first after the participant completed the malfunction resolution. If the participant was unable to resolve the malfunction, the trial was ended at twelve minutes. During the trials, participants' ACAWS display commands (edge key navigations, checklist navigations, switch throws, etc.) were recorded along with their time stamps. Likewise, participants' touches on the PFD were recorded. The participants' eye-movement data were also collected. All trials were video recorded.

Immediately after each trial, questionnaires for NASA Task Load Index (TLX) [6] and modified Bedford workload scales [7] were administered via a computer interface. Following the 7th and 14th trials (i.e., at the end of each session), the participants provided their TLX pairwise comparisons via an electronic questionnaire. After completing both sessions, they provided written answers to questions regarding display usage, display preference, and user interface design.

RESULTS

FAULT MANAGEMENT ACCURACY - The accuracy of each trial was categorized into three classes, *Correct*, *Good*, and *Failed*, based on the following rules. All trials that did not end in the correct final switch configurations were categorized as *Failed*. Also, all trials that did not use the proper checklists for the given situation were categorized as *Failed*. The remaining trials were categorized *Good* if there was any inappropriate switch throw (even if it was corrected later), and *Correct* if all switch throws were correctly done.

The accuracy categorization results by each scenario are listed in Table 2. The total numbers at the bottom indicate that Besi generally had more *Correct* trials (one-tailed binomial t-test, $p = 0.05$) and fewer *Failed* trials (one-tailed binomial t-test, $p = 0.03$) than Elsie. Scenarios #5 through #10 were single-malfunction scenarios, and the majority of participants performed them correctly (*Correct*). These trials will be examined further in the fault-resolution-time and eye-movement-data analyses.

Table 2. Fault Management Accuracy

	Elsie			Besi		
Sc #	Corrct	Good	Failed	Corrct	Good	Failed
1	2	1	1	3	1	0
2	3	0	1	3	1	0
3*	1	0	3	3	0	1
4*	3	0	1	4	0	0
5	4	0	0	4	0	0
6	4	0	0	4	0	0
7	3	0	1	3	0	1
8	4	0	0	3	0	1
9	4	0	0	3	0	1
10	4	0	0	4	0	0
11*	2	0	2	2	2	0
12*	0	2	2	1	0	3
13*	2	0	2	4	0	0
14*	0	2	2	2	2	0
Total	36	5	15	44	5	7

* : Multiple-malfunction scenarios

Scenarios #2 and #3 tested false-alarm cases under single- or multiple-malfunction conditions. Initially, the operator's potential over-trust in Besi's automation was a concern in case HyDE misdiagnosed the root cause. The data, however, did not show any such trend.

Scenarios #11 through #14 were multiple-malfunction scenarios, and generally much harder to resolve than the single-malfunction scenarios. The effects of Elsie and Besi in these multiple-malfunction scenarios were analyzed with a Wilcoxon Signed Ranks Test on the individual participant's performance accuracy categories (*Correct* = 2, *Good* = 1, and *Failed* = 0). For scenarios #11 and #12, there was no significant statistical

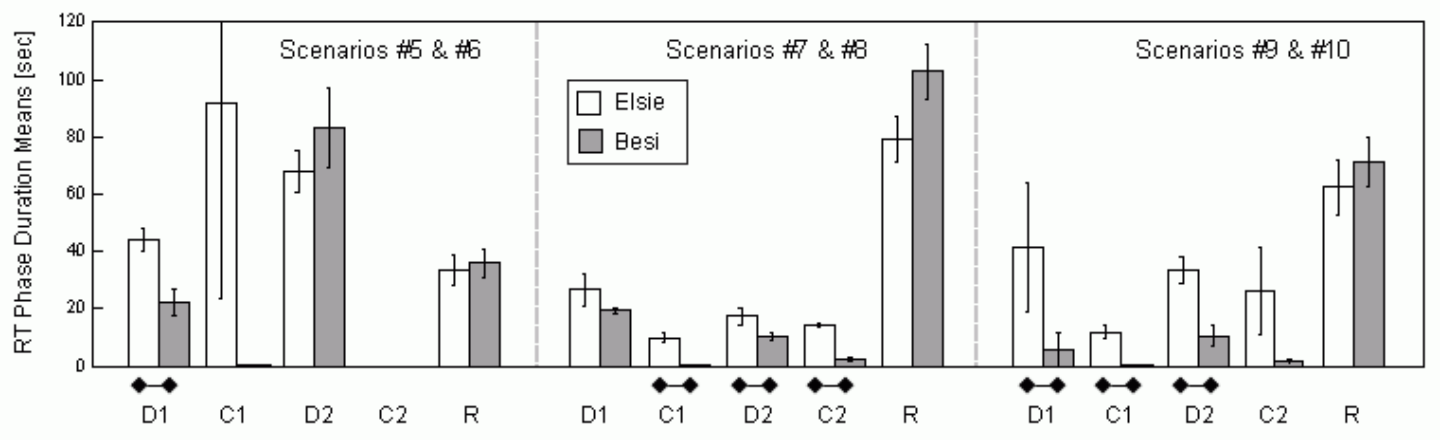


Figure 11: Means and Standard Errors of RT Phase Durations.

difference between the displays. For scenarios #13 and #14, however, Besi performance was significantly more accurate than Elsie ($z = -1.99$, $p = 0.05$).

FAULT RESOLUTION TIME - The fault resolution time (RT) is the time between the first C&W event (i.e., the first fault message) and the time when the last item on the checklist was completed. The RT was computed for scenarios #5 through #10 (the single-malfunction scenarios) because most of the participants performed these procedures correctly, and therefore the RTs were relatively straightforward to compute and fair to compare. (Computing and comparing the RTs became difficult when the participants deviated from standard procedures or interleaved two different procedures. For those reasons, only the single-malfunction trials graded *Correct* were used to compute the RTs.)

Paired t-tests were performed on each participant's RT within each scenario pair (i.e., #5 and #6, #7 and #8, and #9 and #10). Data from participants whose accuracy was categorized as *Failed* in either member of the scenario pair were excluded from this test, so the data were always balanced. The results indicated that, for scenarios #9 and #10, the RTs were significantly shorter for Besi than Elsie ($t(6) = -4.54$, $p < 0.01$, mean RTs = 107.3 sec for Besi, 176.2 sec for Elsie). No statistical significance was found for the other two scenario pairs.

The RT was actually the sum of the durations of the following five phases, each of which consists of different kinds of tasks:

- Diagnosis 1 (D1): the operator assessed the situation and made a root cause determination before selecting the checklist to go to.
- Checklist 1 (C1): the operator navigated to the proper checklist. This process was manual in Elsie, and automated in Besi.
- Diagnosis 2 (D2): the operator followed the checklist instructions to perform the diagnosis steps. Besi's checklists usually contained fewer diagnosis steps than Elsie's.
- Checklist 2 (C2): the operator navigated to the recovery checklist. If there was no separate recovery checklist, C2 duration was zero. Again, going to the recovery checklist was manual in Elsie and automated in Besi.
- Recovery (R): the operator performed the instructed switch throws to reconfigure the system for recovery.

Figure 11 plots the means and standard errors of the RT-phase durations for each scenario pair. Note that the C2 duration was zero for scenarios #5 and #6 because there was no separate recovery checklist. As the figure shows, there were visible trends in the phases D1, C1, D2, and C2, that these durations in Elsie trials were longer than those in the Besi trials except D2 of scenarios #5 and #6. The connected diamonds at the bottom of the bars indicate that the paired t-tests resulted in statistical significance between the durations in Elsie and Besi (for #5 & #6, $t(7) = -18.2$, $p < 0.01$ for D1; for #7 & #8, $t(5) = -5.62$, $p < 0.01$ for C1; $t(5) = -3.47$, $p = 0.02$ for D2, $t(7) = -38.5$, $p < 0.01$ for C2; for #9 & #10, $t(6) = -4.09$, $p < 0.01$ for D1, $t(6) = -5.02$, $p < 0.01$ for C1, and $t(6) = -4.83$, $p < 0.01$ for D2). For the recovery phase (R) durations, the trend reversed; that is, the Besi trials tended to take longer to complete the recovery procedures than Elsie trials, although none of the three scenario pairs achieved statistical significance.

PFD TASK ACCURACY - The PFD task was designed in part to assess the operator's ability to divide attention between malfunction management and monitoring the flight variables. For that reason, only the misses during the time working on any malfunctions (on-task time) were counted for scenarios #5 through #14. In total, 30 color changes on PFD (5.8 % of the total changes) were missed during the Elsie trials, while one color change (0.3%) was missed during the Besi trials. The worst PFD performances were observed during the multiple-malfunction scenarios (#11 through #14), where 24 misses occurred in total in Elsie, and one miss in total in Besi.

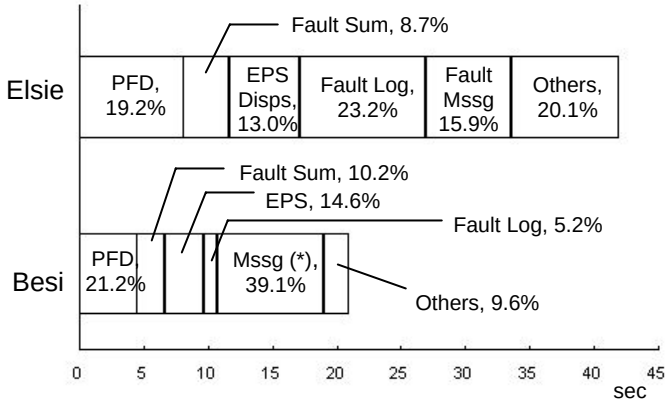
EYE MOVEMENT DATA - The participants' eye-movement data provided a clearer picture of the changes in their PFD scanning strategy. Table 3 summarizes their PFD look statistics during the on-task

time. Within similar difficulty levels (i.e., single or multiple malfunctions), the participants tended to look at the PFD for a higher percentage of the on-task time and with higher frequency when using Besi rather than Elsie. Paired t-test results found significant display effects on them within both the single-malfunction scenarios (scenarios #5 through #10; $t(23) = 3.36$, $p < 0.01$ for the % time; $t(23) = 3.96$, $p < 0.01$ for the number of looks per minute) and the multiple-malfunction scenarios (scenarios #11 through #14; $t(15) = 2.95$, $p = 0.01$ for the % time; $t(15) = 3.24$, $p < 0.01$ for the number of looks per minute). The average PFD look durations did not show any statistically significant ACAWS display effect.

Table 3: PFD look statistics during on-task time

Scenario x ACAWS	% time of PFD look	Number of PFD looks per minute	Average PFD look duration [sec]
Single-mal Elsie	20.6 %	14.6	0.86
Single-mal Besi	26.5 %	17.1	0.95
Multiple-mal Elsie	16.5 %	11.9	0.86
Multiple-mal Besi	20.8 %	14.1	0.90

The participants' display usages during D1 were also computed based on their eye movements. D1 was the only phase, where the participants diagnosed the malfunction mainly based on their free will rather than checklist instructions; and thus, it was especially interesting to see how they used the displays. Figure 12 shows the percentage of the time each display was looked at during the D1 duration of the single-malfunction trials (#5 through #10) for Elsie and Besi. Only the data from the trials graded *Correct* were included. The total lengths of each bar graph are adjusted to be the average D1 durations of Elsie and Besi, 42 and 21 seconds, respectively. The *EPS Disps* category in the Elsie bar graph includes both EPS Sum and EPS Main usages. Both displays present overall view of the EPS status and thus could be used for similar purposes during the initial diagnosis of D1.

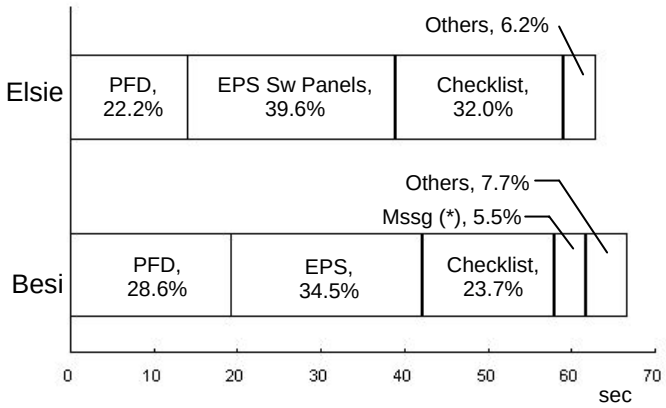


(*): Includes System Status, Root-Cause List, & Fault Mssg areas

Figure 12: Display page usages during D1

The Elsie graph shows that the participants spent the largest amount of time on the Fault Log pages during D1. With Elsie, more than one third of the total D1 time was spent viewing the Fault Log display and the Fault Message area, the two display areas that provided textual information about the system malfunction. When Besi was used, the message part of the display (including system status, root cause list, and fault message areas) was looked at for more than one third of the total D1 time. Note this area showed the root cause diagnosis results and was also used to switch the displays in the main area. These may be the reasons why this part of Besi received many fixations during D1.

In addition, the participants' display usages during R were also calculated to examine the reason why the R durations were longer in Besi than in Elsie (see Figure 11). Figure 13 shows the percentage of the time each display was looked at during the R phases of the single-malfunction trials (#5 through #10) for Elsie and Besi. Only the data from the trials graded *Correct* were included in the graph. The total lengths of each bar graph are adjusted to be the average R durations of Elsie and Besi, 63 and 67 seconds, respectively. The graphs show that the participants looked at PFD for longer in Besi than in Elsie.



(*): Includes System Status, Root-Cause List, & Fault Mssg areas

Figure 13: Display page usages during R

To further analyze the changes in their PFD-monitoring strategies during R, paired t-tests were applied to the PFD-look statistics during R in the single-malfunction trials (#5 through #10). The results indicated that the participants looked at the PFD for significantly larger percentages of the R durations with Besi than with Elsie ($t(22) = 2.29$, $p = 0.03$), and the average PFD look durations were marginally longer with Besi than with Elsie ($t(22) = 1.86$, $p = 0.077$). No statistical significance was found for the number of PFD looks per minute. Note one participant did not finish the R phase of scenario #7 in Besi, and thus, this trial and the paired trial (#8 in Elsie) of this participant were not included in the t-tests.

SUBJECTIVE WORKLOAD - The TLX workload scores and the Bedford workload scores of the paired scenarios (#5 through #14) were tested with a Generalized Linear Model (GLM). The main effects included were Participant, Display (Elsie vs. Besi), and Malfunction (Single vs. Multi), and the interaction effect included was Display x Malfunction. The results showed that the TLX scores were significantly lower (less workload) in the Besi trials than in the Elsie trials ($F(1,69) = 36.40$, $p < 0.01$, TLX mean = 3.77 for Besi, 5.34 for Elsie), and also significantly lower in single-malfunction scenarios ($F(1,69) = 44.57$, $p < 0.01$, TLX mean = 3.80 for Single, 5.69 for Multi). The Participant effect was also significant in the TLX scores ($F(7,69) = 14.27$, $p < 0.01$). The interaction effect also turned out to be significant ($F(1,69) = 9.51$, $p = 0.02$); That means that the reduction of the TLX scores by Besi compared to Elsie was greater in the multiple-malfunction trials than in single-malfunction trials. See Figure 14.

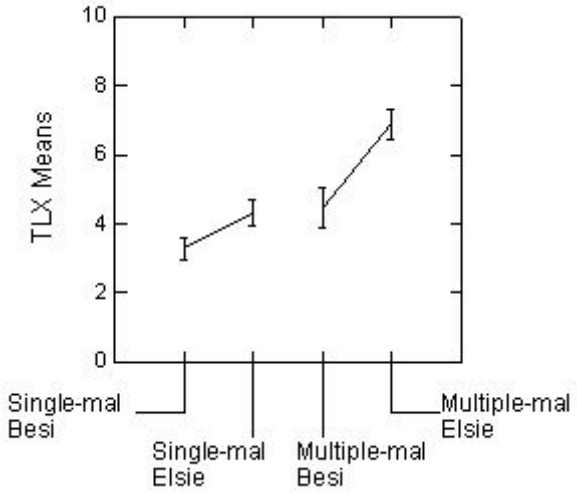


Figure 14: Grand means of TLX scores

The same GLM analysis was applied to the modified Bedford workload scores. The results showed that the Bedford scores were significantly lower (less workload) in the Besi trials than in the Elsie trials ($F(1,69) = 3.06$, $p < 0.01$, Bedford score mean = 2.75 for Besi, 3.88 for Elsie), and in single-malfunction trials than in the multiple-malfunction trials ($F(1,69) = 30.39$, $p < 0.01$, Bedford score mean = 2.56 for Single, 4.44 for Multi). The Participant effect was also significant in the Bedford scores ($F(7,69) = 11.05$, $p < 0.01$). Unlike in the TLX, no significant interaction effect was found in the Bedford scores.

SUBJECTIVE PREFERENCE AND USABILITY - The participants answered the display preference questionnaire after completing all 14 trials. The questionnaire presented continuous scales for display preference scores. The left end of the scale was "Absolutely Preferred Elsie," and the right end was "Absolutely Preferred Besi." The mid point was "Neutral." One-sample t-test results showed that the participants preferred Besi significantly over Elsie for the diagnostic processes ($t(7) = 7.35$, $p < 0.01$), recovery processes ($t(7) = 4.47$, $p < 0.01$), and overall processes as well ($t(7) = 4.69$, $p < 0.01$).

The questionnaire also asked their preference between Elsie's textual representations of EPS parameters vs. Besi's graphical representations of EPS parameters. A one-sample t-test result showed that they marginally preferred Besi's graphical representation over Elsie's textual representation ($t(7) = 2.27$, $p = 0.058$). The questionnaire also asked their preference between Elsie's switch panels presented separately from the EPS line diagram vs. Besi's switch indicators integrated into the EPS line diagram. A one-sample t-test result indicated that they significantly preferred Besi's integrated switch indicators over Elsie's separate switch panels ($t(7) = 2.31$, $p = 0.05$).

DISCUSSION

The study results generally proved that Besi has many advantages over Elsie. According to the preference questionnaire, our participants overwhelmingly preferred Besi over Elsie in diagnosis, recovery, and overall processes. The participants felt that the workload was significantly lower when they were using Besi than Elsie. The performance data also showed that the participants could correctly perform the malfunction management procedures more frequently with Besi than Elsie. Especially during the scenario pair #13 and #14 (double battery failures), the participants benefited from Besi's advantages and more participants could complete the complex procedures correctly with Besi than with Elsie. During the single-malfunction scenarios (#5 through #10), the procedures were straightforward enough that most participants could perform them correctly with both Elsie and Besi, and thus the benefit may have not been obvious from the accuracy data alone. However, the differences were in the PFD task performance and the RTs.

The PFD task performance data showed that the participants missed the color changes much more frequently with Elsie than with Besi during both single-malfunction scenarios (six misses to zero) and multiple-malfunction scenarios (24 to one). That means Elsie caused attention tunneling more frequently than Besi. The eye-movement data also supported the finding. When the participants were using Besi, they looked at the PFD for a higher percentage of the time and with a greater sampling rate than when they were using Elsie. During an actual space flight operation, it is critical for the crew to always keep a close eye on the flight variables even during system malfunctions. Therefore, some may say this finding is the most significant indicator of the benefits of Besi.

Besi also shortened the durations of the RT phases, D1, C1, D2, and C2 (See Figure 11). The results were not surprising. D1 was shorter with Besi than with Elsie because Besi's automated root-cause diagnosis

capability eliminated the need to read through the fault messages. C1 was shorter with Besi because of Besi's function of automatically bringing up the appropriate checklist, eliminating the operator's need to manually navigate to the checklist. D2 was shorter with Besi because the automated root-cause diagnosis software already performed the basic parameter cross-checking, and therefore the operator needed to perform only a minimum verification step instead of the lengthy diagnosis steps that Elsie typically requires. C2 was shorter with Besi because it, again, automatically brought up the appropriate recovery checklist.

Even though the results themselves were not surprising, these capabilities of Besi may provide useful hints for the designers, should it be decided to not install the automated root-cause diagnosis software due to other factors, such as vehicle weight. Even if the automated root-cause diagnosis capability is not available onboard, for instance, the Fault Log could be designed so it directly links to the checklist, and the checklist corresponding to a particular fault message could be brought up instantly from the Fault Log. This would shorten or eliminate the C1 time. Likewise, if a checklist line requires jumping to another checklist, the line can be directly hyperlinked to the proper checklist so that it pops open automatically. The computer could also automatically perform basic parameter cross-checking to shorten the D2 time.

Moreover, the eye-movement data results showed that, when using Elsie, the participants spent over one third of D1 time reading the fault messages. The current Fault Log and Fault Message (similar to those used in the Space Shuttle) present the fault messages in chronological order. Instead, if these messages could be shown in a way that reflects the system hierarchy information, so that finding the message from the most upstream component becomes easier, that could accelerate the fault diagnosis process in D1.

One place that Besi did not result in shorter durations was the R phase. In Besi, the system reconfiguration process tended to take as long as or longer than in Elsie. The eye-movement analyses revealed that the participants monitored the PFD for larger percentages of the R durations in Besi than in Elsie. In fact, the usages of the other displays (e.g., EPS, Checklist) were actually slightly shorter with Besi than with Elsie, but the participants looked at the PFD long enough in Besi to the extent that it eventually made the Besi's overall R durations longer than those of Elsie. During the R phase, the larger percentage of the PFD look was caused by slightly longer average duration of each PFD look, rather than increased PFD-visit frequencies. The participants may have felt that they had extra time to look at the PFD when they were using Besi.

The longer R durations in Besi may have been caused by the participants themselves rather than the Besi's display design, but it may be still possible to shorten the Besi's R durations by improving the display design. As mentioned before, Besi's EPS display included switch symbols directly controllable within the line diagram. This resulted in many selectable features crowding within a single diagram. Getting to the switches was still easy because the *Connect* edge key transferred the cursor directly to the switch associated with the current line. However, getting the cursor out from the diagram required manually moving the cursor to the *Return to Edge Keys* icon located at the upper-right corner of the main area (see Figure 9). Because of the many selectable features in between, the process sometimes took long. To shorten the path, the cursor movements were designed to be circular at each edge of the display: For instance, if the cursor was at the lower-left corner of the EPS display, pressing the *Down* arrow key and the *Left* arrow key moved it to the upper-right corner of the EPS display. However, it was often observed that the participants failed to utilize this short cut, and ended up taking the longer path directly toward the icon. One way to accelerate the switch-throwing processes, and in turn to shorten the R durations in Besi, is to equip the hand controller with a button that let the cursor out of the diagram by one press.

One of Besi's design principles was to present summarized information to reduce display clutter. Besi had the *View* edge key at the upper left corner to reveal the suppressed information (i.e., the numerical readings of voltage, current, etc., and inactive connections), but none used that feature during the on-task time in this study. They may have felt little need to look up the suppressed information, or may have been too busy to do so, or both. It is important to note that the task in this study could be completed without the information suppressed in Besi. Thus, the study's results should not be considered supporting *any* information suppression. (For instance, if the tasks required reading the numerical values of certain parameters, the current Besi design may have been inappropriate.) Rather, the results mean that information suppression should be considered carefully with the operator's task in mind. The results did show that having an option to look up extra information does not necessarily mean the operators actually voluntarily look it up, especially under high workload situation.

Finally, the participants' possible over-trust on Besi's automated root-cause diagnosis software was initially a concern. The study did not find any hard evidence of such case, but it is possible that the scenarios used in the study were too simple. This is such a critical issue that further research should be conducted before concluding. HyDE, the automated root-cause diagnosis software, worked well in the current study (except the sensor failure cases, where it was forced to misdiagnose), but no software is perfect. Also, the longer the duration of space operation becomes (such as the trip to Mars), the more chance that the internal model of the diagnosis software will deviate from the actual spacecraft states. As a result, the automated root-cause diagnosis software might become unreliable at some point. The automation could be very beneficial in

accelerating the fault-management processes and reducing the operator workload, as found in the current study's results. However, the ACAWS displays should be also designed so that the operator can easily recognize any incoherent response of the automated root-cause diagnosis and override the fault management procedure when necessary.

CONCLUSION

Two potential concepts of CEV cockpit ACAWS displays for EPS and ECLSS fault management - Elsie and Besi - were examined. The results of a human-in-the-loop simulator experiment with eight instrument-rated aircraft pilot participants showed that more participants could solve a complex multiple-malfunction case correctly with Besi than with Elsie. Also, in the single-malfunction cases, Besi's automated root-cause diagnosis capability and the direct link from the root-cause message to the corresponding checklist significantly shortened the time for diagnosing the malfunction and navigating to the appropriate checklist. In one of the single-malfunction scenarios, Besi trials showed 40% reduction of the total fault resolution time compared to Elsie trials. The participants monitored the flight variables on the PFD more often when they were working on fault management with Besi than with Elsie. The questionnaire results showed that the participants felt less workload when they were using Besi than Elsie, and also the participants overwhelmingly preferred Besi over Elsie for both diagnosis and recovery processes. The participants strongly preferred Besi's switch symbol representations integrated in the system line-diagram, compared to Elsie's virtual switch panel representations, and marginally preferred Besi's graphical representation over Elsie's textual representation. The study did not find any hard evidence of the participants' over-trust of Besi's automation, but the scenarios used here may have been too simple. Further study should be conducted regarding automation trust issues.

ACKNOWLEDGMENTS

This work was supported by the Space Human Factors and Habitability Element (WBS 466199) of NASA's Human Research Program, the ARC Crew Cabin 1% ESMD AUG (WBS 64423.02.35.02.99), the ARMD Aviation Safety Program (IVHM Project), and the ARMD NRA, Aviation Safety: Integrated Vehicle Health Management Project, Solicitation #NNH06ZEA001N-IVHM. We thank Scott Poll, Adam Sweet, William Taylor, John Ossenfort, Valerie Huemer, Steven Elkins, David Garcia, Matthew Daigle, Indranil Roychoudhury, Ann Patterson-Hine, Serdar Uckun, David Nishikawa, David Hall, Charles Lee, Christian Neukom, Serge Yentus, and Lee Stone for their hardware and software support. Last but not least, we thank all of our pilot participants for their patience and support.

ACRONYMS

ACAWS: Advanced Caution and Warning System
C&W: Caution and Warning
C1: Checklist 1 (second of the five RT phases)
C2: Checklist 2 (forth of the five RT phases)
CAU: Cockpit Avionics Upgrade
CEV: Crew Exploration Vehicle
D1: Diagnosis 1 (first of the five RT phases)
D2: Diagnosis 2 (third of the five RT phases)
ECLSS: Environmental Control and Life Support System
EPS: Electrical Power System
GLM: Generalized Linear Model
HyDE: Hybrid Diagnostic Engine
NASA: National Aeronautics and Space Administration
PFD: Primary Flight Display
TLX: (NASA) Task Load Index
R: Recovery (fifth of the five RT phases)
RT: Resolution Time

REFERENCES

1. McCandless, J. W., R. S. McCann & B. R. Hilty, "Upgrades to the Caution and Warning System of the Space Shuttle," Human Factors and Ergonomics Society 47th Annual Meeting, Oct. 13-17, Denver, CO, 2003.
2. Maluf, D. A., Y. O. Gawdiak & D. G. Bell, "On Space Exploration and Human Error - a Paper on Reliability and Safety," the 38th Annual Hawaii International Conference on System Sciences, Jan. 3-6, Waikoloa, HI, 2005.
3. Hayashi, M., V. Huemer, F. Renema, S. Elkins, J. W. McCandless & R. S. McCann, "Effects of the Space Shuttle Cockpit Avionics Upgrade on Crewmember Performance and Situation Awareness," Human Factors and Ergonomics Society 49th Annual Meeting, Sep. 26-30, Orlando, FL, 2005.
4. Wickens, C. D., S. E. Gordon & Y. Liu, An Introduction to Human Factors Engineering, Addison-Wesley Educational Publications Inc., ISBN: 0321012291, New York City, NY, 1998.
5. Narasimhan, S. & L. Brownston, "Hyde - a General Framework for Stochastic and Hybrid Model-Based Diagnosis," the 18th International Workshop on the Principles of Diagnosis, May 29-31, Nashville, TN, 2007.
6. Hart, S. G. & L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in Hancock, P. A. & N. Meshkati (eds.), Human Mental Workload, North Holland Press, Amsterdam, 1988.
7. Huntley, M. S., "Flight Technical Error for Category B Non-Precision Approaches and Missed Approaches Using Non-Differential Gps for Course Guidance," US Dept. of Transportation Volpe National Transportation Systems Center, DOT-VNTSC-FAA-93-17, Cambridge, MA, 1993.